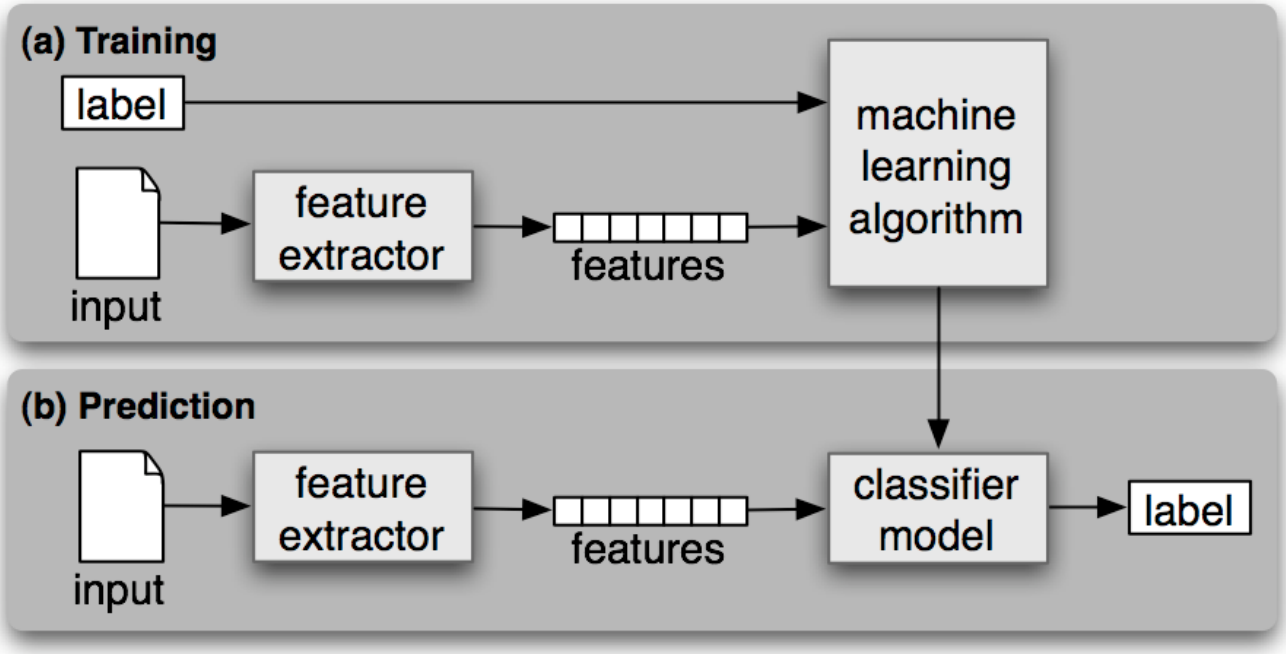
Computerlinguistische Anwendungen
Support Vector Machines

Thang Vu

CIS, LMU
thangvu@cis.uni-muenchen.de

May 20, 2015

Introduction



- Today we will learn about a new classification algorithm: Support Vector Machines (SVMs)

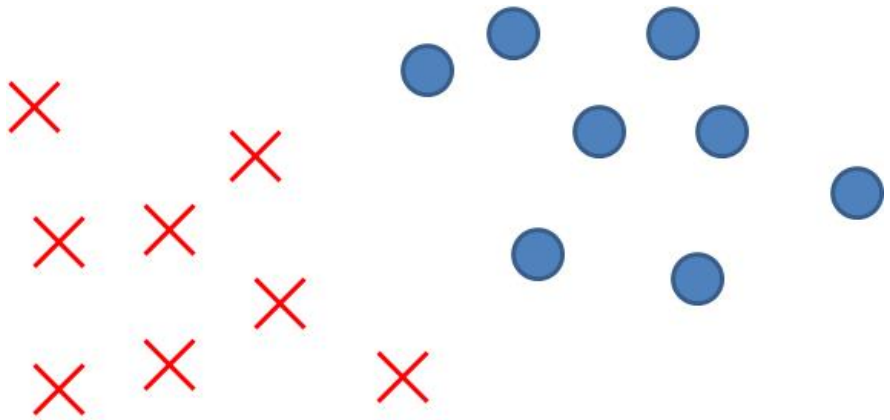
Introduction

- The original SVM algorithm was invented by Vapnik V. and Chervoneskis A. in 1963
- It was extended and improved for machine learning applications, published in 1995 by Cortes C. and Vapnik V.
- Raised interest of a big number of researchers and users
- It proved to be useful in many machine learning applications such as computer vision, text classification, etc.
- Still state-of-the-art technique for many natural language processing applications

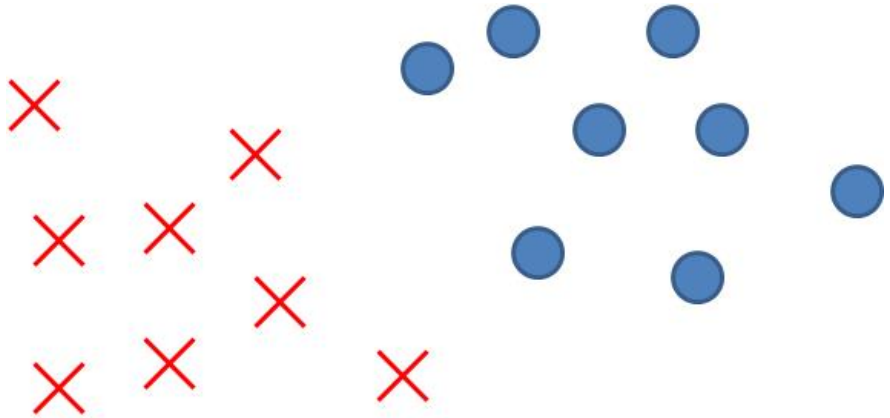
Introduction

- SVMs are supervised learning methods
  $\rightarrow$ Need labeled data
- Can be used for **classification** and regression
- It works not only in linear classification but also non-linear classification

Classification: linear separable case

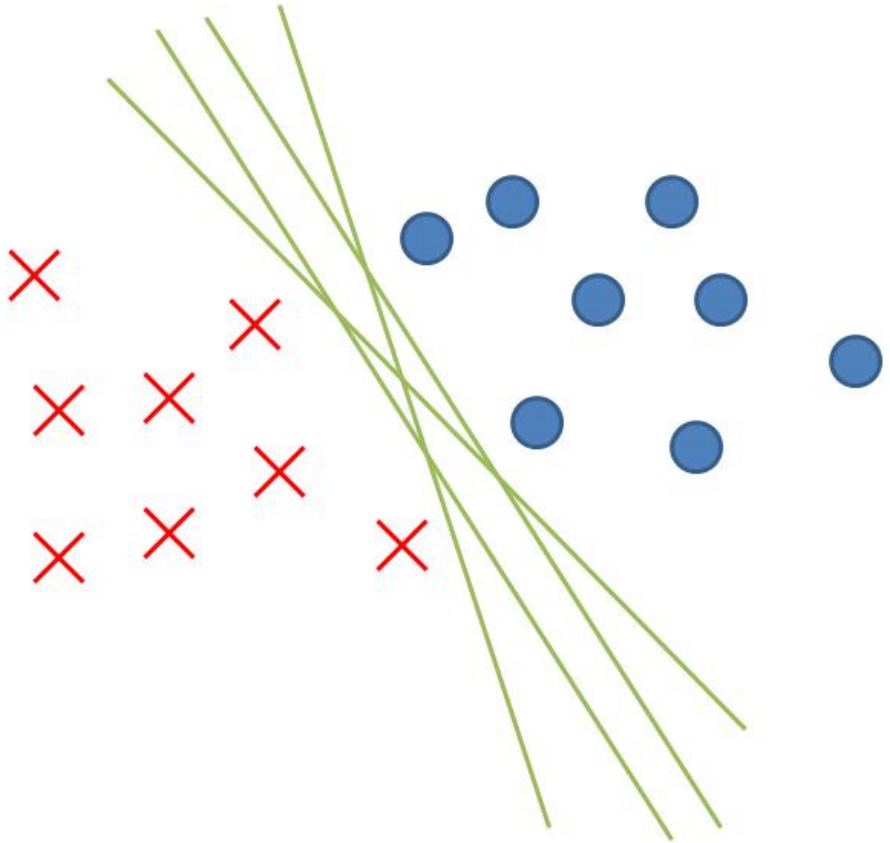


Classification: linear separable case



- Where is the best linear separator?

Support vector machine: motivation

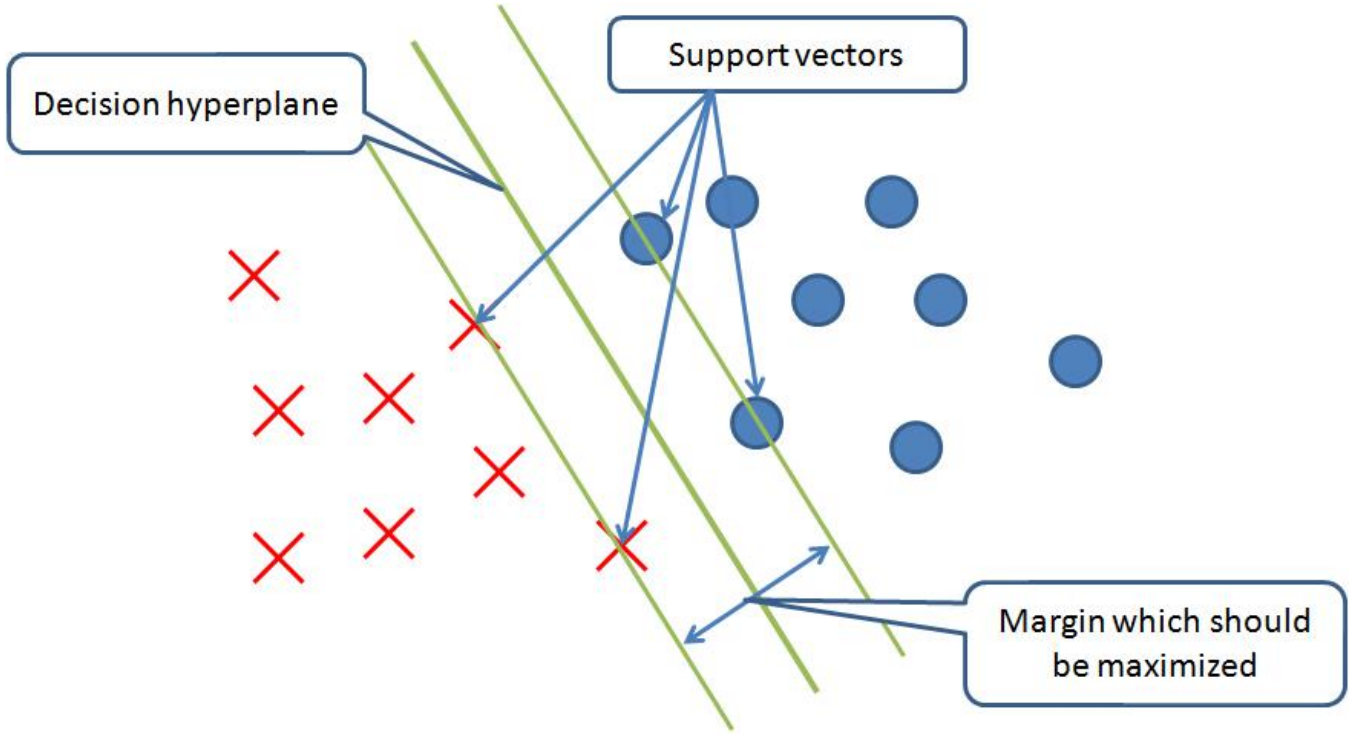


- Where is the best linear separator?

Support vector machine: Idea

- SVMs looks for a decision surface that is maximally far away from any data point
- The distance from the decision surface to the closet data point is the **margin** of the classifier
- Maximizing the **margin** improves the generalization of the model

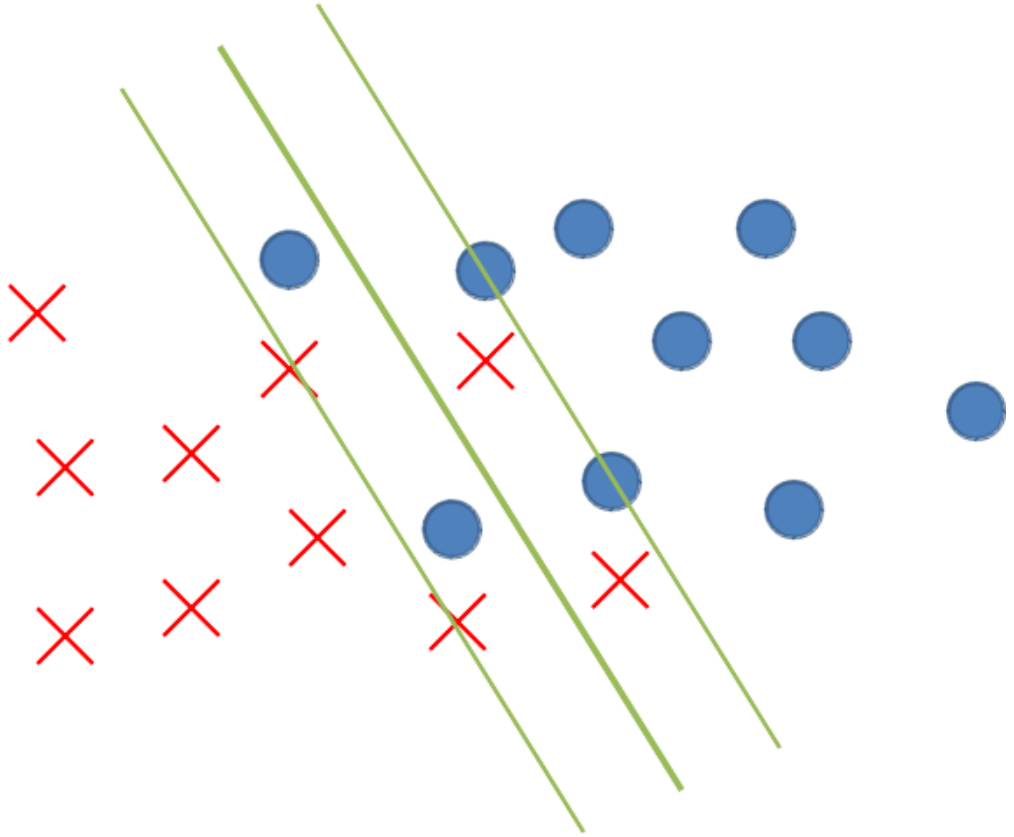
Support vector machine



Support vectors

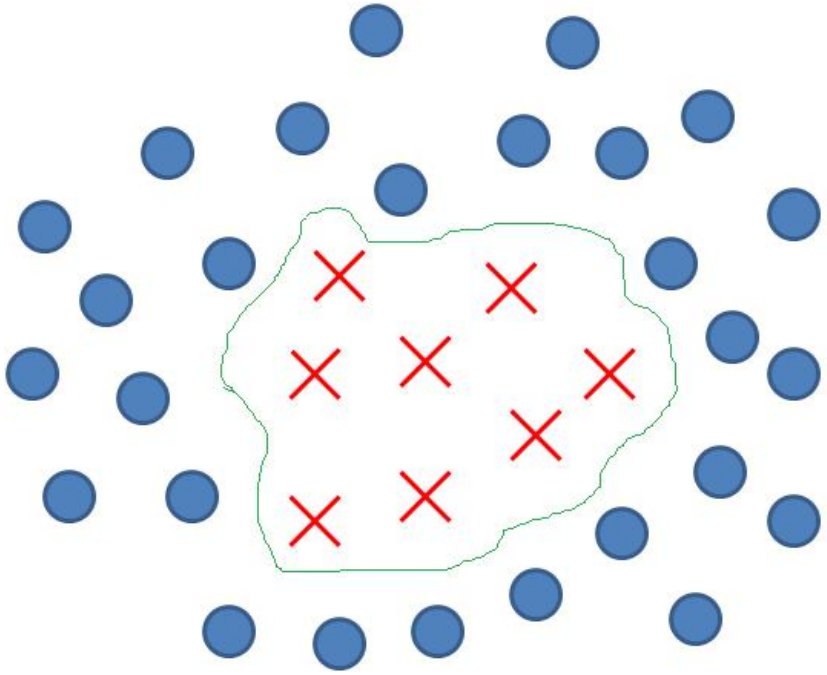
- Linear hyperplane is defined based on the support vectors
- Changing other points a bit or adding more training data which are already well classified might not help
- Only need to store support vectors to predict the labels of the new points

Not linear separable case

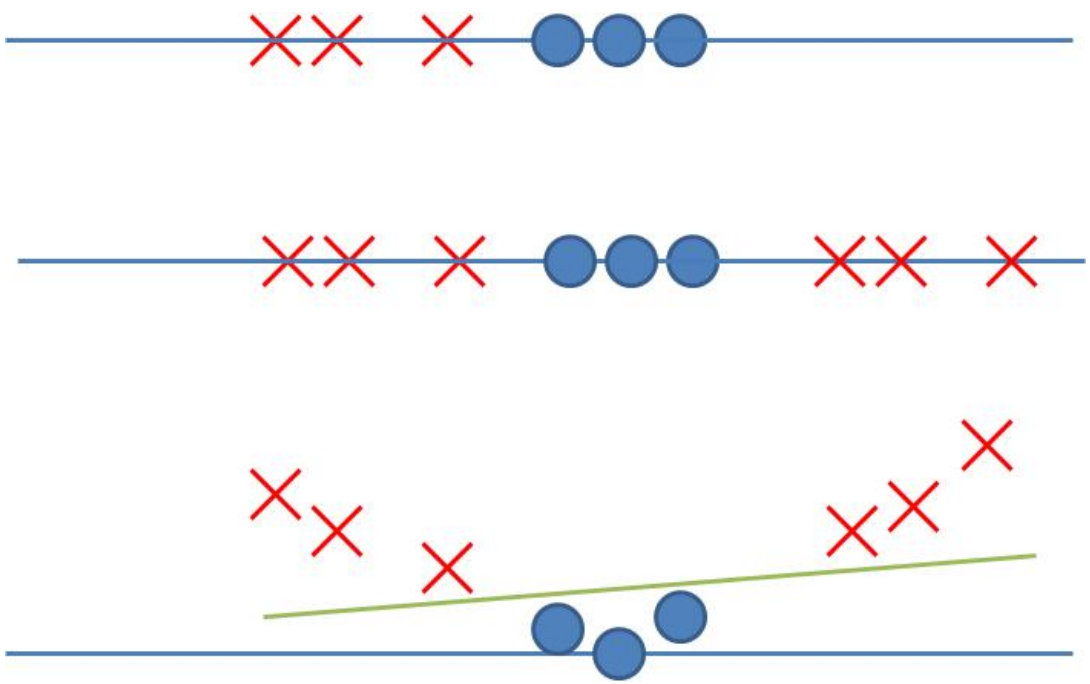


- Allow error during training
- Maximize the margin and minimize the error on the training data

Classification: non-linear separable case



Kernel SVM: intuition



Kernel SVM

- Mapping data points to a higher dimensional space $x \rightarrow \psi(x)$
- In this space, data might be linear separable
  i.e. $\rightarrow$ Non-linear problem $\rightarrow$ linear problem
- It is referred to as “the kernel trick ”

Support vector machine with Scikitlearn

- There are 3 different SVMs implementations in Scikitlearn: SVC, NuSVC, LinearSVC

```
1 #ABC is the name of the SVMs
2 from sklearn.svm import ABC
3 #(X_train, y_train) = (features matrix, labels)
4 #X_test = features matrix for testing
5
6 model = ABC()
7 model.fit(X_train, y_train)
8
9 y_predict = model.predict(X_test)
```

SVC

- Time complexity is more than quadratic with the number of samples $\rightarrow$ It is not suitable for > 10k samples
- Multiclass classification: one-vs-one scheme
- Important parameters:
  - C
  - gamma for rbf and poly kernel
  - coef0 for poly and sigmoid kernel
  - etc.

```
1 from sklearn.svm import SVC
2 model = SVC()
```

LinearSVC

- Similar to SVC with parameter kernel = 'linear'
- However, it has more flexibility in the choice of parameters
- It runs faster especially with large number of samples
- Multiclass classification: one-vs-rest scheme
- Important parameters:
  - C
  - etc.
- It was used in many NLP research papers

```
1 from sklearn.svm import LinearSVC
2 model = LinearSVC()
```